TITLE: DIFFERENCE BETWEEN BIOINFORMATICS PIPELINES FOR THE ANALYSES OF SOIL BACTERIAL MICROBIOTA USING 16S rDNA GENE SEQUENCER DATA.

AUTHORS: CAVALCANTE, J.P.C.M.; SILVA, L.C.F.; LIMA, H.S.; SILVA, C.C.;

INSTITUTION: UNIVERSIDADE FEDERAL DE VIÇOSA - UFV, VIÇOSA-MG, MG (AV. PH ROLFS, SEM NÚMERO, CEP 36570000, VIÇOSA – MG, BRAZIL).

ABSTRACT: In the last ten years, the access for soil microbial diversity was improved with popularization of new technologies by Next Generation Sequence like Pirosequence 454/Roche and Illumina platform. This popularization makes possible some advances in microbial ecology, however each sequencing generates a huge sequencing data that needs sophisticated analyses for bioinformatics pipelines. Nowadays, there are some bioinformatics pipelines that are different but can using the same database, so, the results of both analysis should be similar. The aim of this study is to compare two pipelines, one internet-based and other that using command line protocols for relative abundance and diversity index. Twenty-four sampling of soil was obtained in the central area of Minas Gerais state, Brazil, and the total DNA of the samples was extracted with Nucleospin soil Kit according to the manufacturer's recommendations. The DNA obtained was amplified for 16S rDNA and sequencing by Illumina MiSeq. The data were analysed with MG-RAST (Metagenomics Rapid Annotation using Subsystems tecnhnology) and Qiime (Quantitative insights into microbial ecology) using Brazilian Microbiome project pipeline following patters steps for each pipelines. For both pipelines the database used was SILVA, diversity index and statistics analyses were performed with PAST. The results were different between pipelines. At phyla level for all samplings, Qiime/BMP show Proteobacteria as main group with 46,1% of relative abundance followed for Actinobacteria with 20,2%, curiously in MG-rast the main group was Actinobacteria (42%) followed for Proteobacteria (31%). The diversity index of Shannon in all samples by Qiime/BMP was higher than the same data for MG-RAST at OTU Level with 97% of similarity. The results show significant difference between pipelines, at phyla level, the order of main group were changed since both pipelines used the same dataset and database, this difference may be due to the algorithm used by each program. Thus, it is concluded that the choice of pipeline and platform to be used in the analyses is an important step, since different choices can lead to different results. Perhaps analysis of data on more than one platform should be recommended or establish some criteria that are important to be evaluated in each pipeline.

Keywords: Bacterial Diversity; Soil Microbiome; Metagenomic; Gene Survey; Illumina MiSeq.